



europeana
think culture

'Semantic' Libraries: For Which Purpose?

Prof. Dr. Stefan Gradmann
Humboldt-Universität zu Berlin / School of Library and Information Science
stefan.gradmann@ibi.hu-berlin.de



Overview



From the Web of Documents to the Linked Data Web

... others will cover

From Containers to Content to Context

... opportunities for libraries after the erosion of the 'book' at the end of the Gutenberg Paranthesis

... what for? For Whom?

Digital Scholarship and DM2E



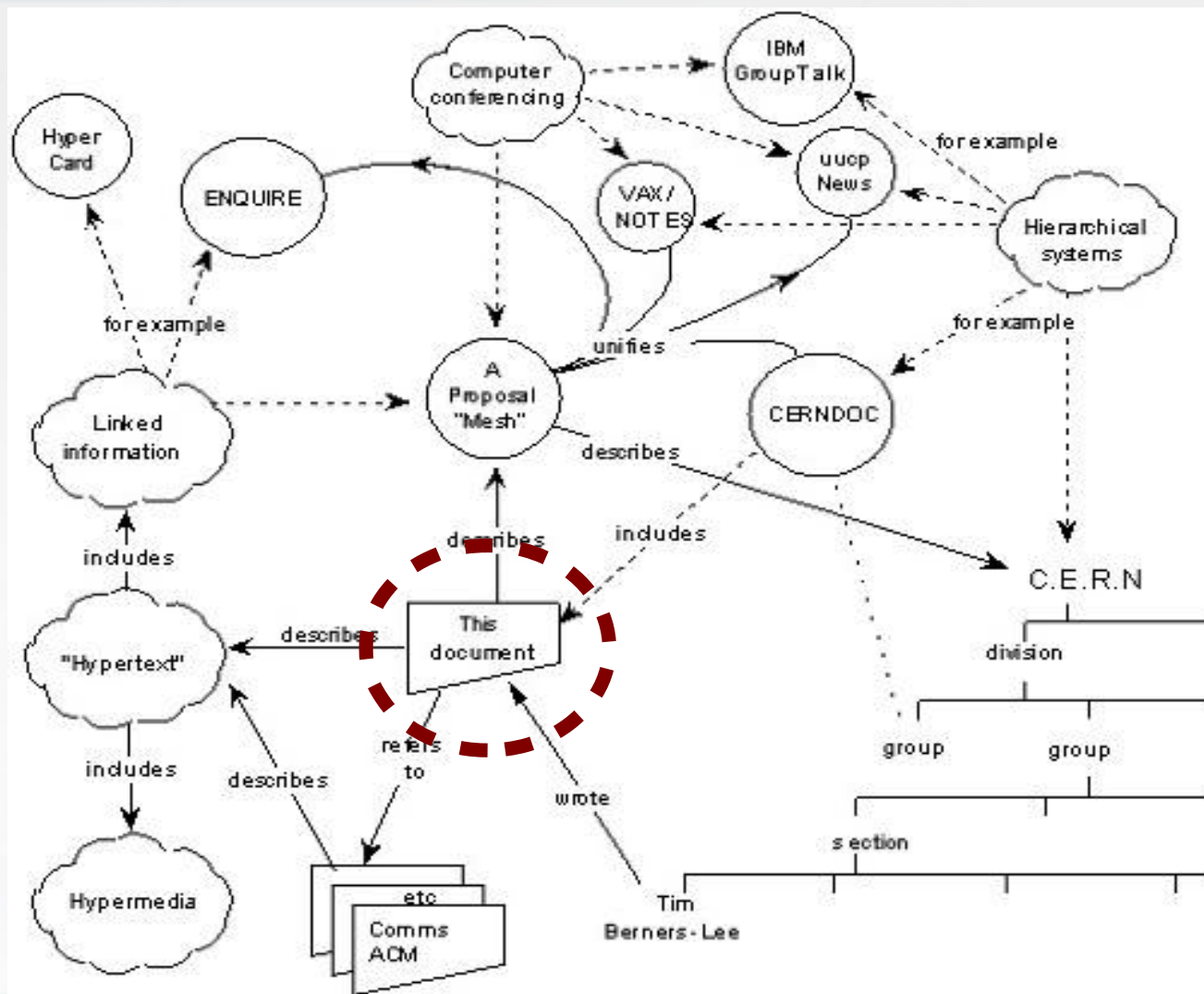
From the Web of 'Documents' to the 'Linked Data Web'

The talk I expect others to give

(or which you can find at <http://www.slideshare.net/gradmans/20120424-container-contentcontextbielefeldmini>)



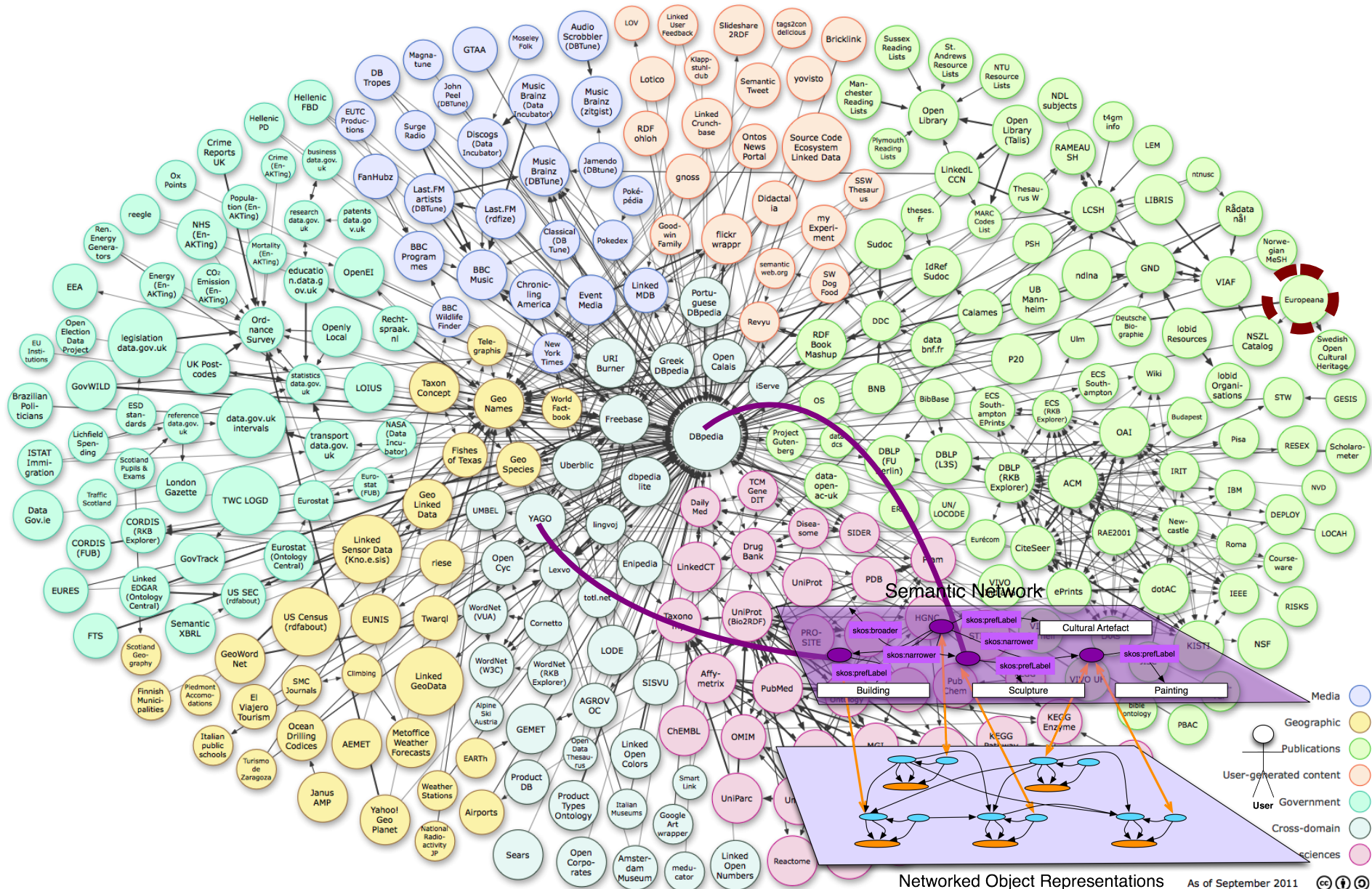
The Web of Documents



Information Management:
A Proposal
(TBL, 1989)

How we've
extended
this model
in **syntax**
and **scope**

EDM and Linked Open Data



SEMANTIC LIBRARIES: FOR WHICH PURPOSE?

Stefan Gradmann, Madrid / Berlin, June 18 2012

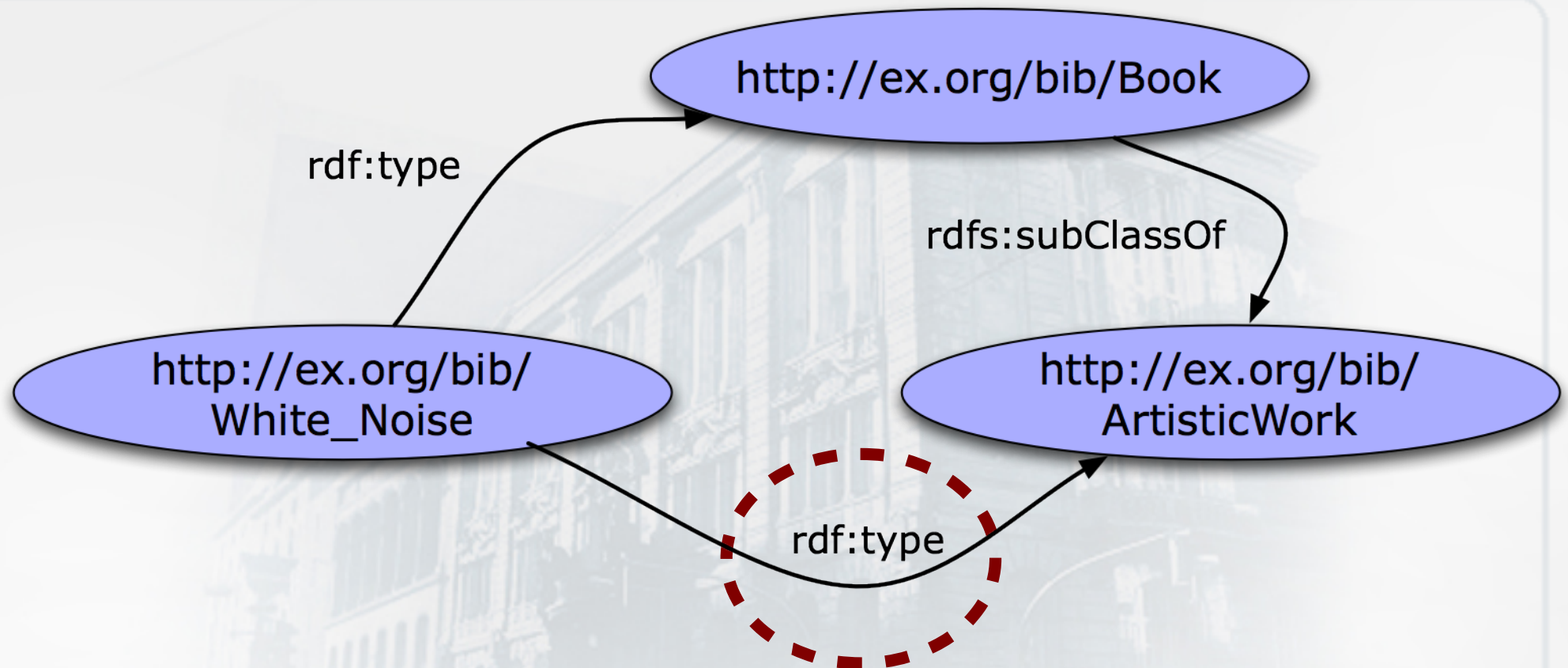
Machines can reason on triple sets!



Some reasoning preconditions ...



... and an automated inference!



There is quite some potential for scholarly scenarios here!

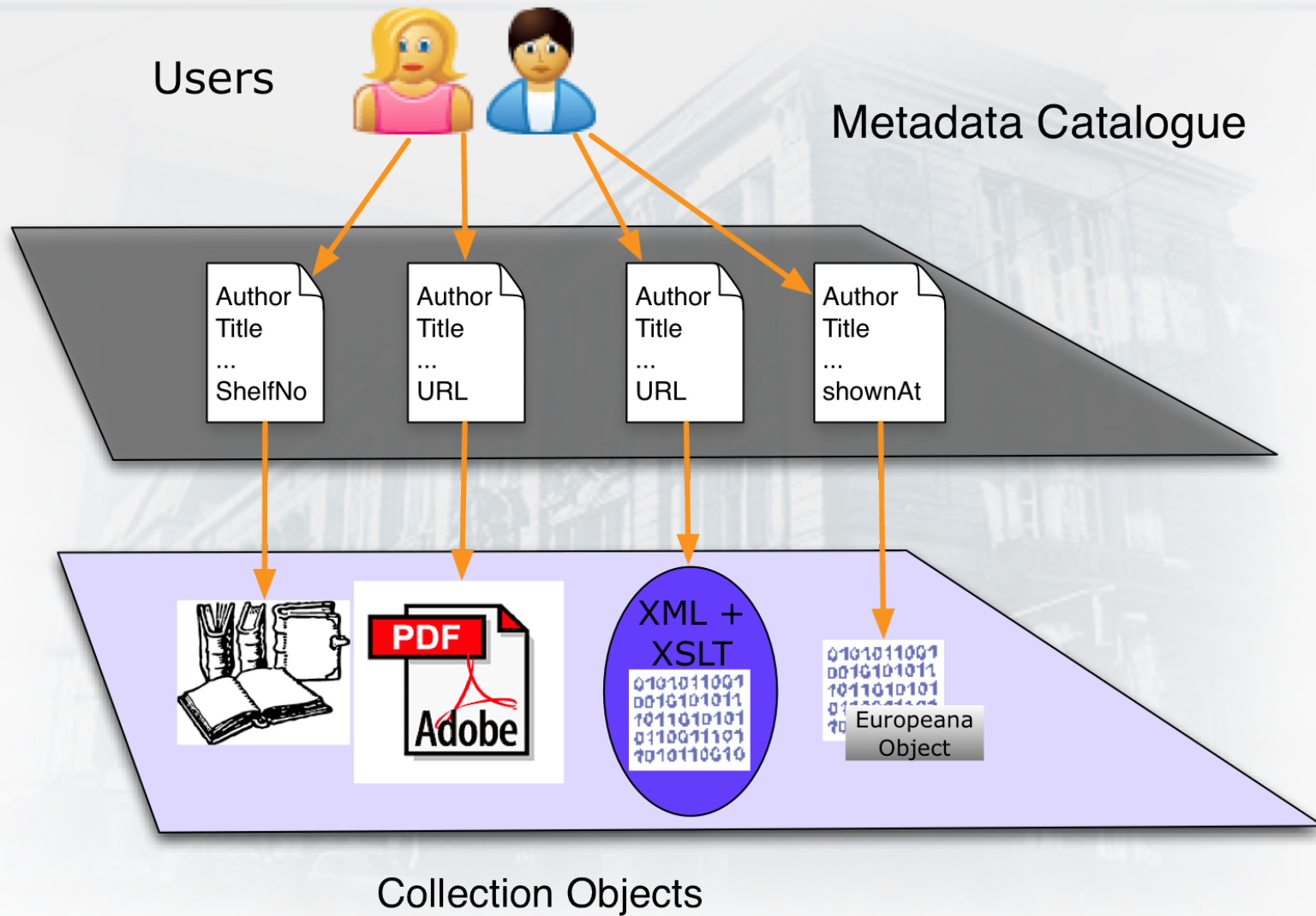


An Opportunity for Libraries ...

... and what they need to change
to be up to it



Catalogues and Information Containers



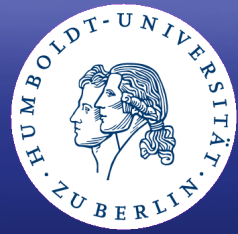
Library Functional Principles (2)



- **Mediating access** to information objects via **catalogues**
- **Mediating links** as pointers from metadata to objects
- Objects are part of a library **collection**
 - An object to be used within a library typically is part of this library's collection
- Internal processing logic: focus on
 - 'bookish' objects as information **containers**,
 - not so much on the **content** of these containers
 - and accordingly **cataloguing** is focussed on **container attributes**
- All this is falling to pieces with the gradual **disintegration of 'monolithic' container formats.**



“What do you do with a million books?” (G. Crane)



- Digitisation and semantic publishing result in
 - growing quantity
 - increased complexity
- Well beyond scholarly processing capacity (=reading faculty)
- Multiplication of collections or distributors is annoying
→ as few as possible. Ideally just one (?)
- Scientists and Scholars will badly need help in :
 - **Semantic abstracting, named entity recognition** for “strategic reading” (Renear)
 - **Contextualisation** of information objects
 - Robust **reasoning and inferencing** yielding digital heuristics
- => ***Potential opportunities for libraries ...***



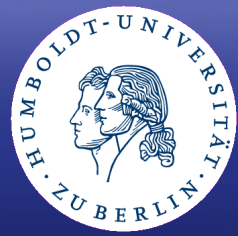
Ceci n'est pas une bibliothèque



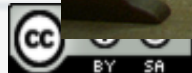
Ceci n'est pas une bibliothèque



Catalogue



The card catalog in the nave of Sterling Memorial Library at Yale University. Picture by Henry Trotter, 2005.



Catalogue Entry: MARC Record



LEADER: 00495cam 2200169 a 4500
001 94205672
003 DLC
005 20040108112243.0
008 940907s1926 gw 000 1 ger
010 \$a 94205672
040 \$aDLC\$cDLC\$dDLC
050 00 \$aPT2621.A26\$bS36 1926
100 1 \$aKafka, Franz,\$d1883-1924
245 14 \$aDas Schloss :\$bRoman /\$cFranz Kafka.
260 \$aMünchen :\$bKurt Wolff Verlag,\$cc1926.
300 \$a503 p. ;\$c20 cm.
561 \$aSource: Bequest of Aaron Copland, Fall, 1989.\$5DLC

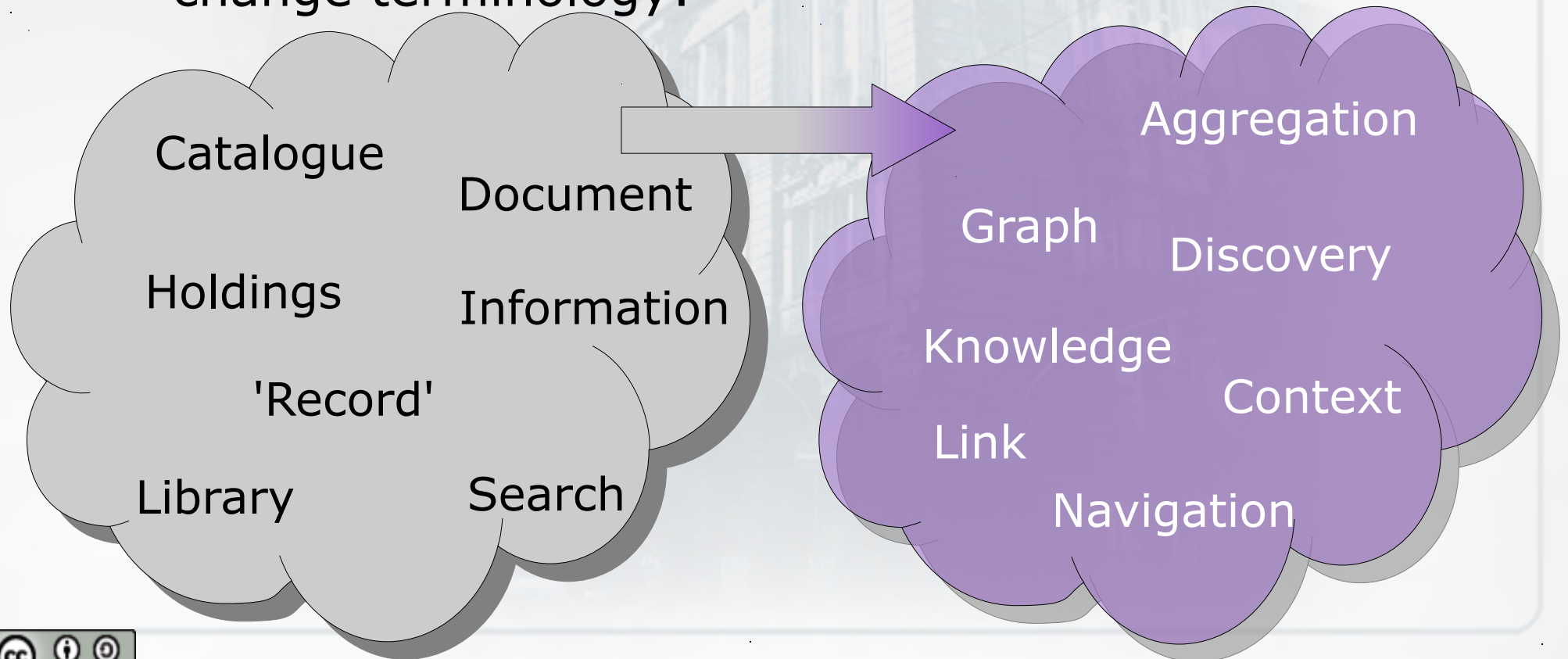


'Library Collections'

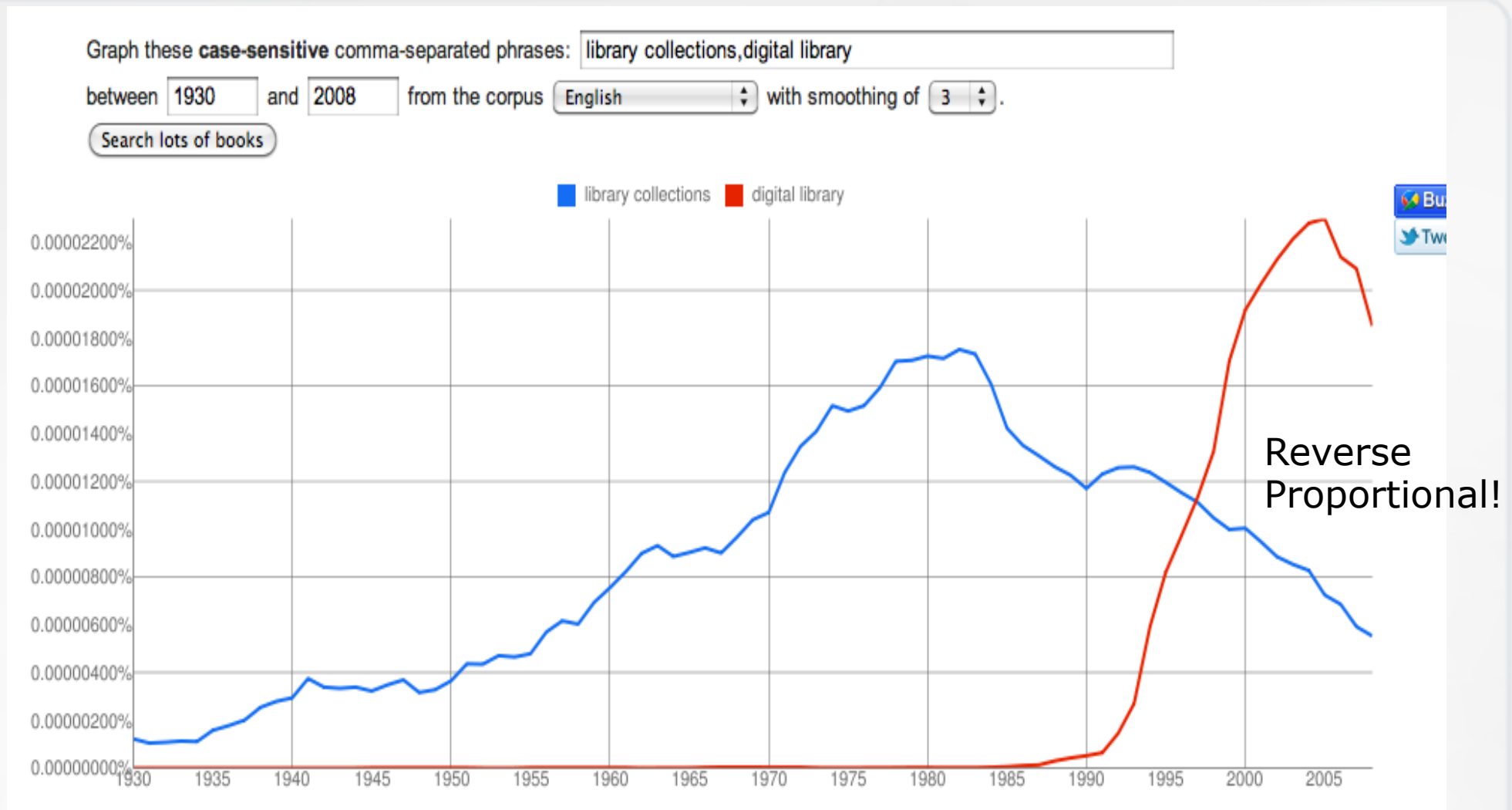
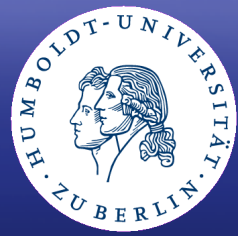


Change Terminology!

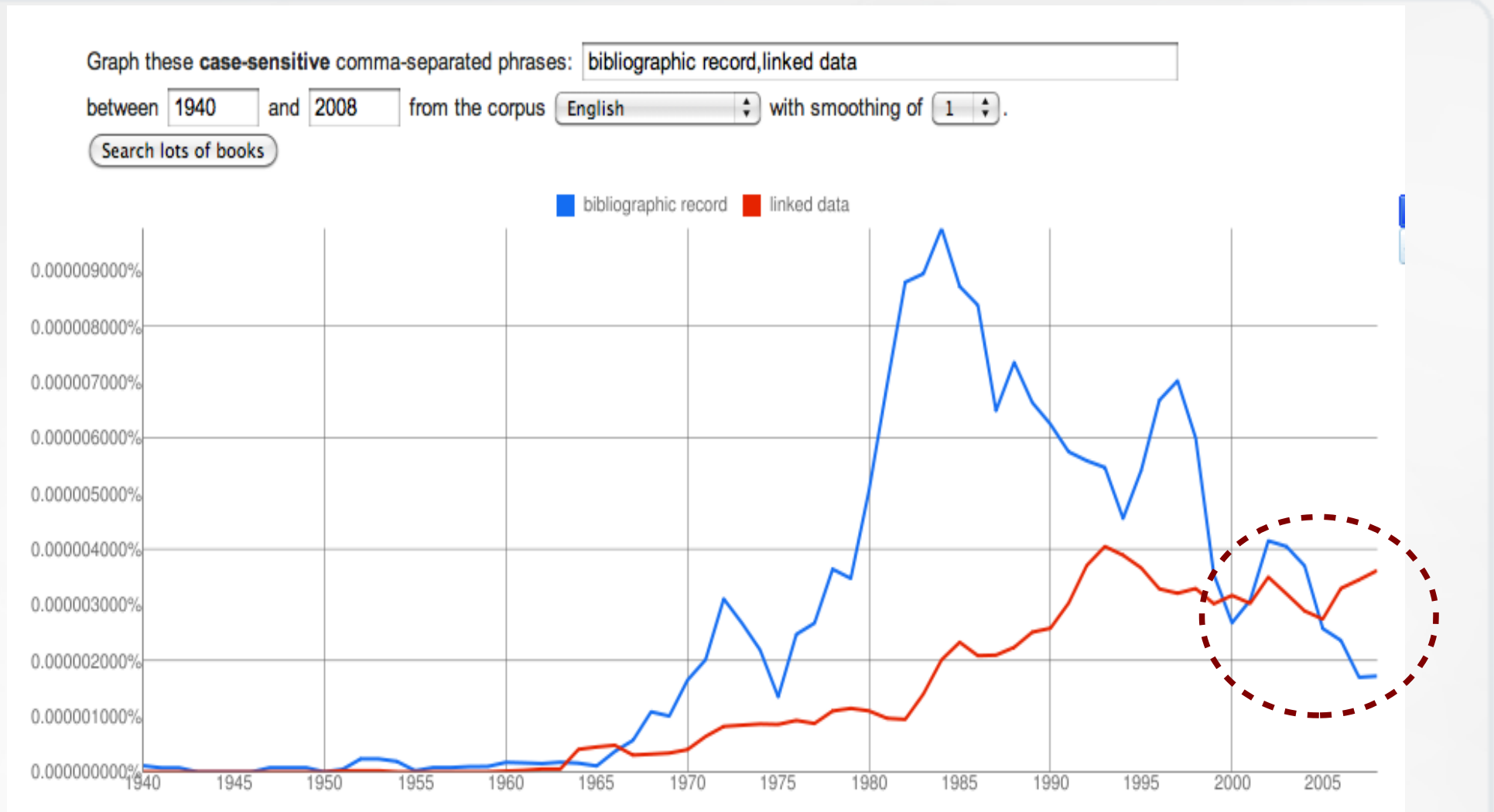
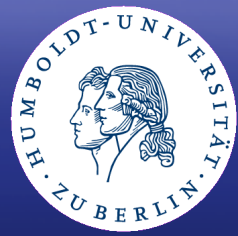
- Libraries will serve research as part of the Linked Open Data web – or else risk becoming insignificant.
- For operating this change they definitely need to change terminology:



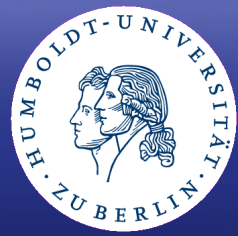
From 'Catalogues' to 'Graphs': old terms – new terms (1)



From 'Catalogues' to 'Graphs': old terms – new terms (2)



From 'Catalogues' to 'Graphs': old terms – new terms (3)

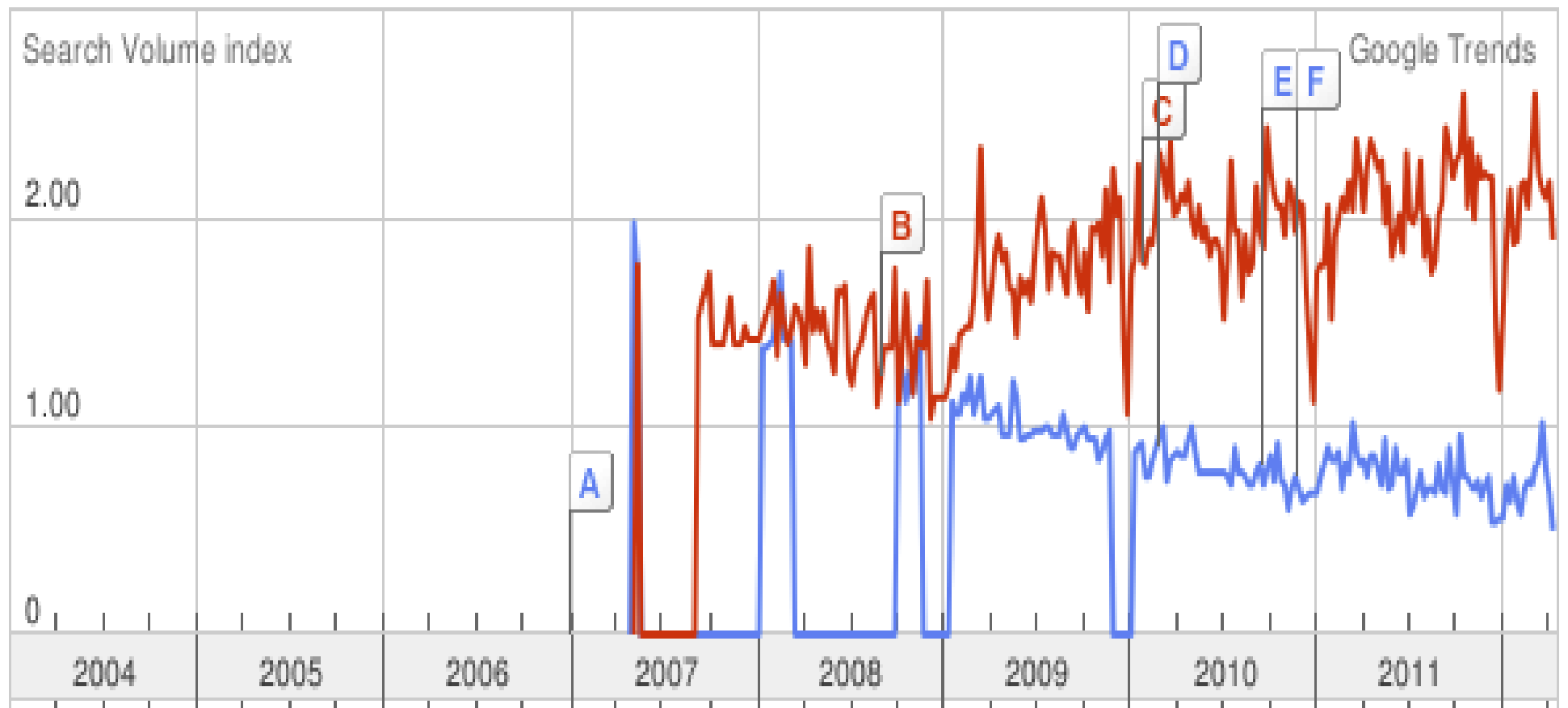


cataloguing

0.38

linked data

1.00



Lessons learned in Europeana

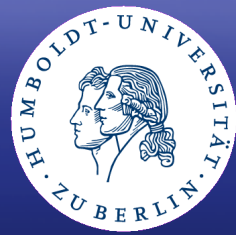


- We have learned some of these lessons in Europeana
 - we dropped the brand "EDL" very early
 - we decided not to have a 'catalogue'
- We know that the current portal is not enough
 - we devised the RDF based Europeana Data Model (EDM)
 - we are gradually migrating to EDM based operations
 - we make Europeana part of the Linked Open Data cloud

What for?



Digitised Manuscripts to Europeana (DM2E): Who (1)?



■ Content Providers

- European Association for Jewish Culture (Judaica)
- Max-Planck-Institut für Wissenschaftsgeschichte (ECHO)
- Österreichische Nationalbibliothek (Google)
- Staatsbibliothek zu Berlin (Kalliope)
- University of Bergen (Wittgenstein)
- CNRS ITEM (Nietzsche)
- National Library of Israel (Judaica)
- Berlin Brandenburgische Akademie (German Text Archive)
- Humboldt-Universität zu Berlin (Polytechnisches Journal)

■ Technology Providers

- ExLibris (Aleph, MARC sources management)
- Freie Universität Berlin (dBpedia, LoD, LoD2, D2R, SILK)
- Max-Planck-Institut für Wissenschaftsgeschichte (ECHO)
- Net7 S.r.l. (Muruca/Pundit)
- National Technical University of Athens (MINT)



Digitised Manuscripts to Europeana (DM2E): Who (2)?



■ **Digital Humanities Community**

- Dr. Tobias Blanke (King's College, London)
- Sally Chambers (The European Library / DARIAH-D)
- Prof. Dr. Stefan Gradmann (HUB)
- Prof. Dr. Gerhard Lauer (Göttingen University)
- Dr. Alois Pichler (UIB)
- Dr. Jürgen Renn (MPIWG)
- Dr. Laurent Romary (HUB)
- Prof. Dr. Susan Schreibman (Trinity College Dublin)
- Dr. Claire Warwick (University College, London)

■ **Community Building**

- Open Knowledge Foundation (OKFN)

■ **Coordination, Management & Information Science**

- Humboldt-Universität zu Berlin (HUB)

■ **TEL** (Europeana Research)



Digitised Manuscripts to Europeana (DM2E): What?



- WP1: Provide substantial amounts of **digital content** to Europeana with a focus on **digitised manuscripts**
- WP2: Integrate existing **technical building blocks**
 - from Europeana development
 - as well as from generic LoD oriented development
 - into a **generic production chain** for migrating data from various sources to the EDM as well
 - as for the **contextualisation** of the object representations.
- WP3: Explore usage scenarios of such content in a specialised platform for **humanities research** generating digital heuristics and making data as well as heuristics available to specialised visualisation or reasoning environments.



WP3: Digital Humanities Requirements and Related Engineering - Context



Goal: lower the barriers for **digital content curation** by providing an integrated, flexible, semantic based environment targeted to **digital humanities scholars**

Structured or semi-structured data sources

Compliant web pages with metadata

Europeana (EDM)

Partners/third party platforms

Linked Open Data Cloud

Web of Data

Digital re-mix and curation tool
for humanities scholars and domain experts

Text collation

Multimedia (semantic) annotation

Text mining

Data Linking

(personal) digital curation workflows

Mix, recombine digital content and metadata

Augment knowledge based on Linked Data

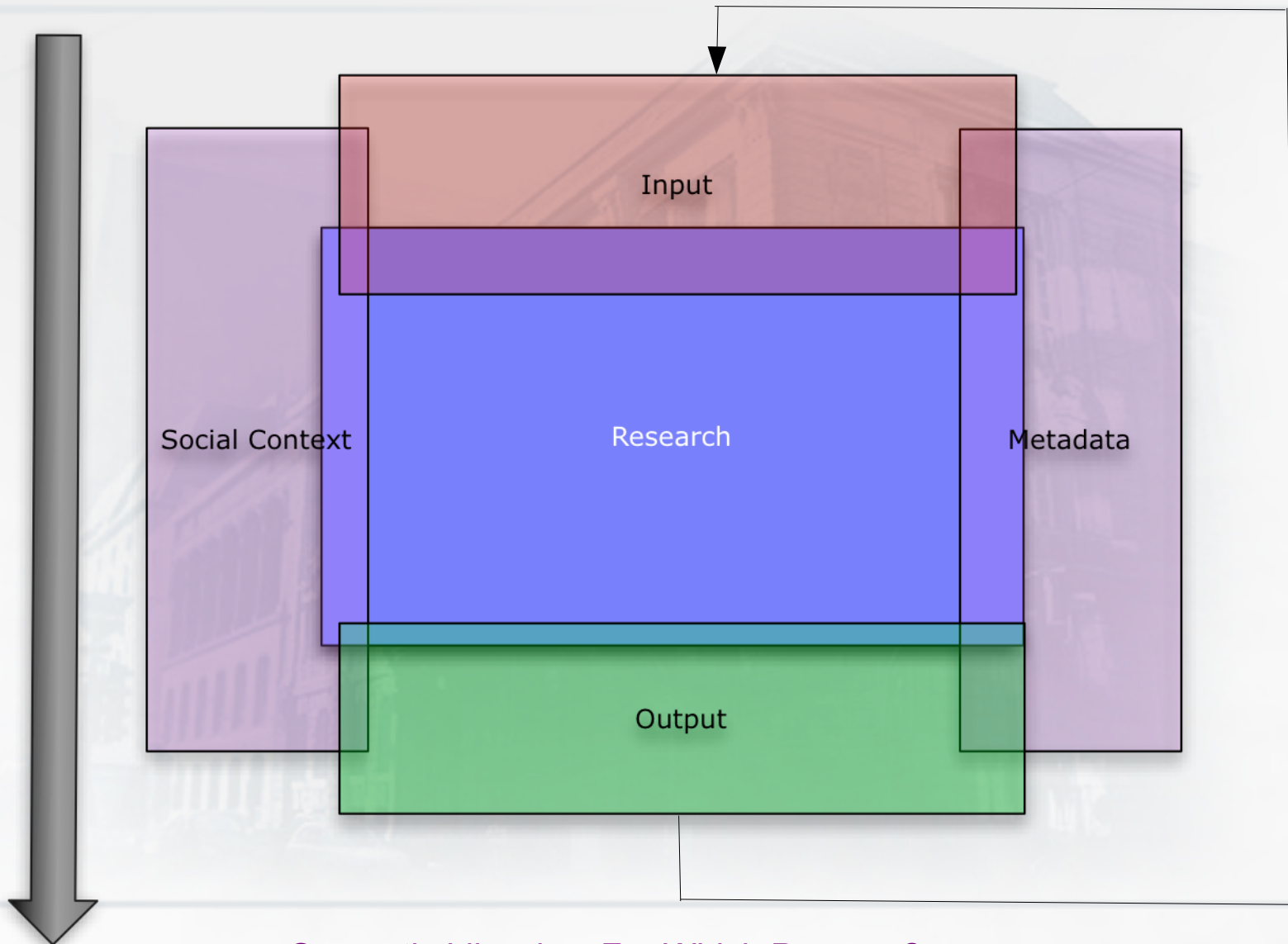
- Custom virtual collections
- Curated, domain specific mashups
- Scientific contributions
- etc..

Scholarly Communities

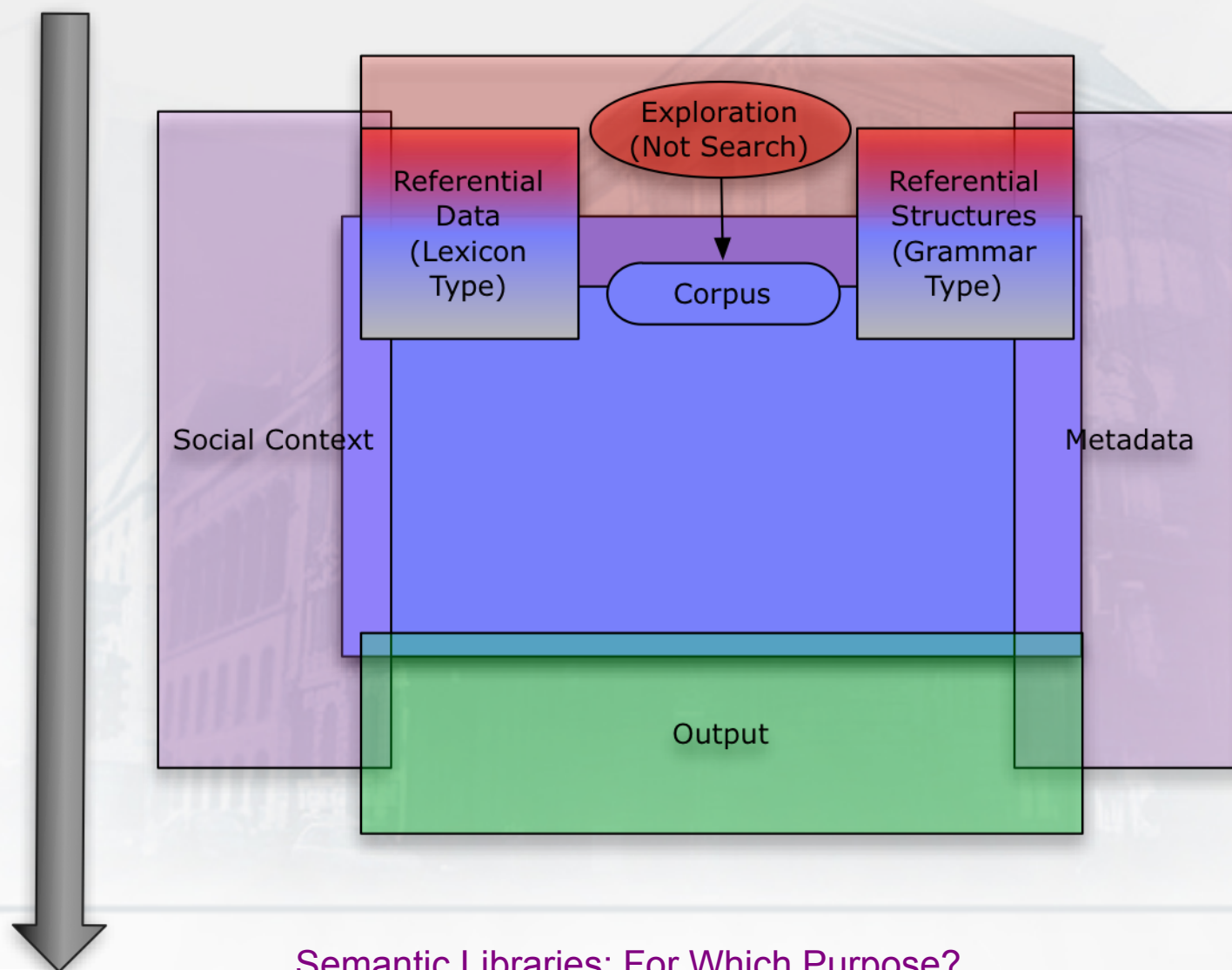
Linked, semantically structured content
further enriching the data web



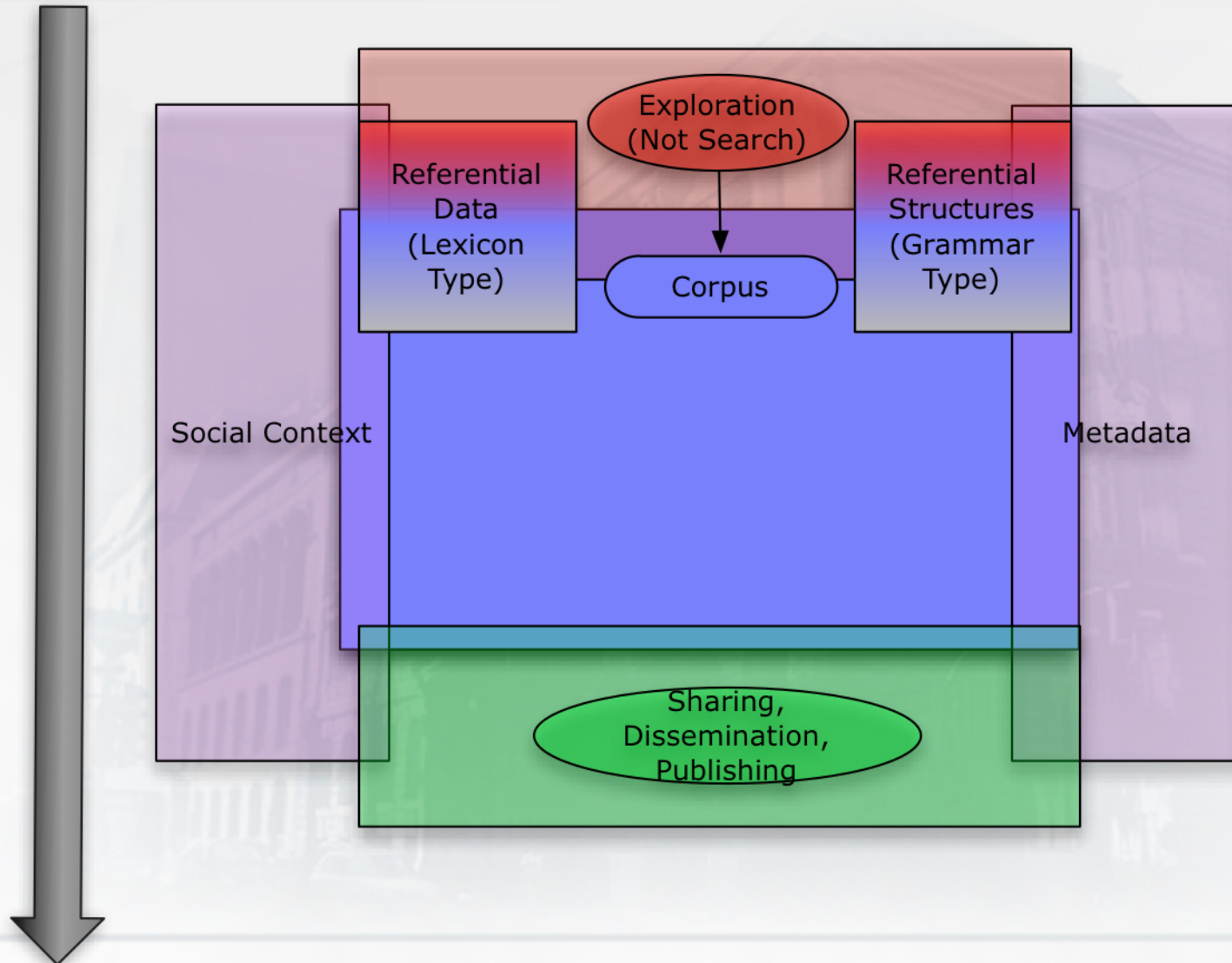
The Scholarly Domain from 10.000 feet above



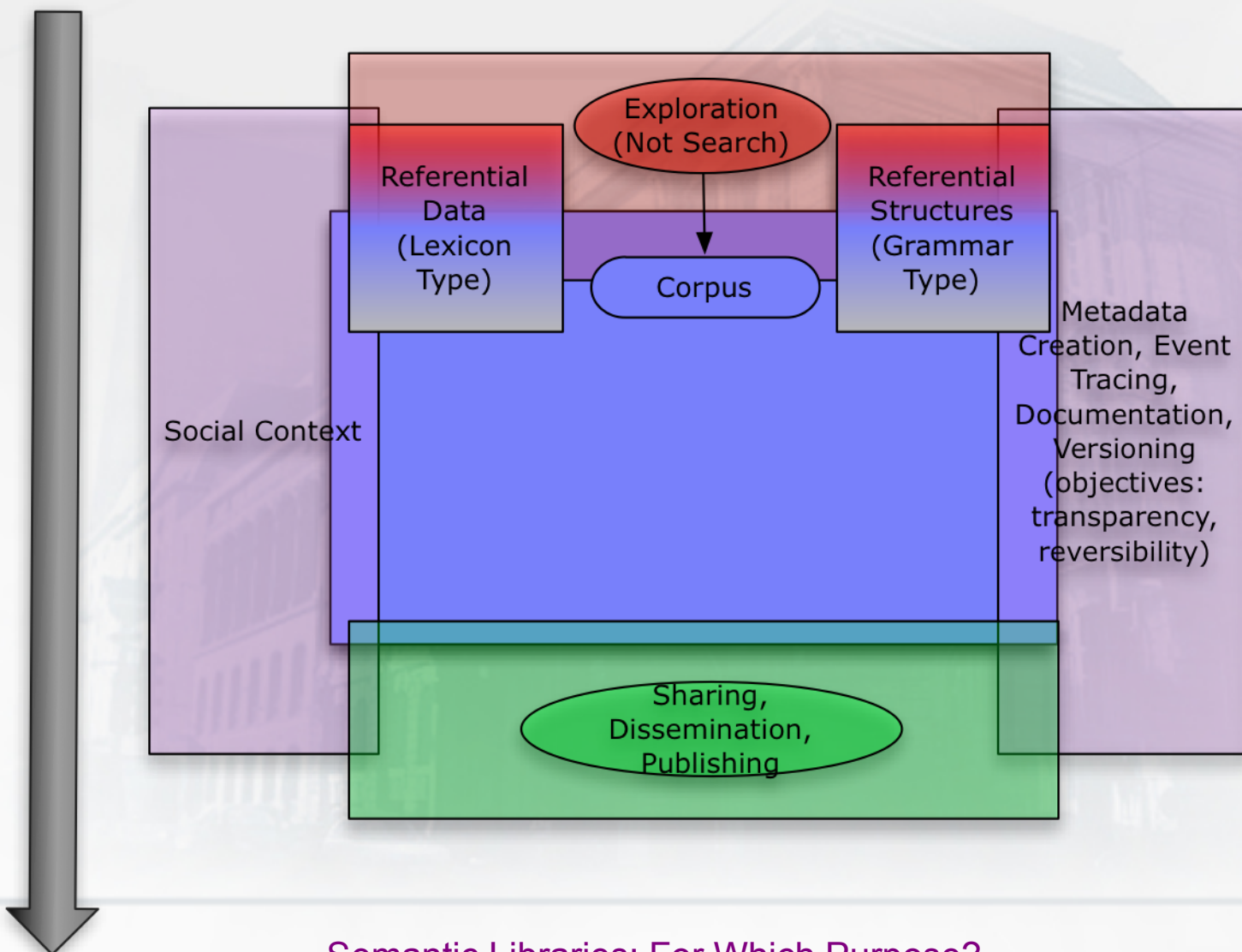
Input Area Details



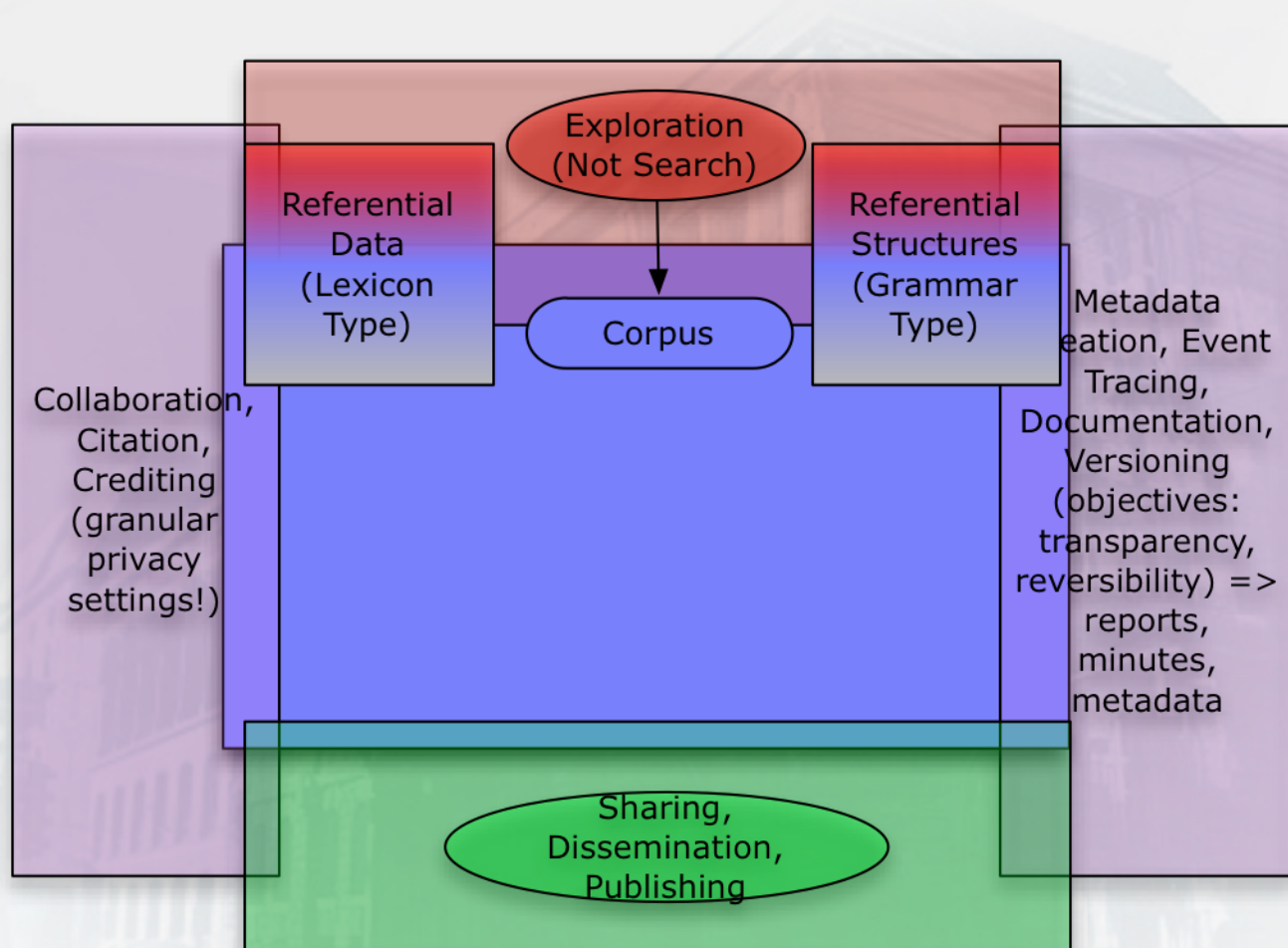
+ Output



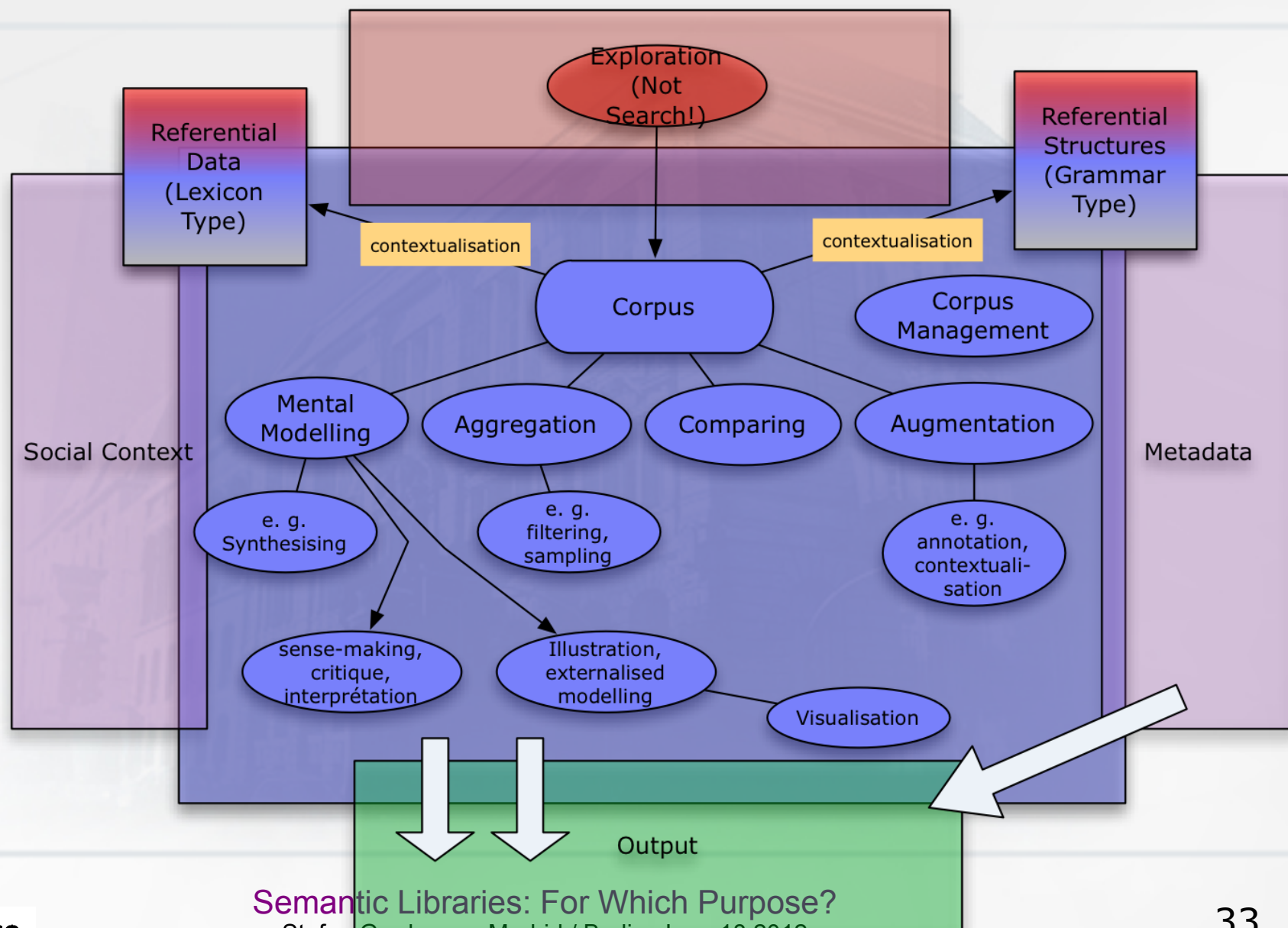
+ Metadata



+ Social Context



+ Zoom on Research



Semantic Libraries: For Which Purpose?

Stefan Gradmann, Madrid / Berlin, June 18 2012

Roadmap

- Identify the intended functional extension of the 1st Pundit & Korbo versions
- Stabilise scholarly domain model
- Identify additional specialisations of primitives
- Formalise, ontologically model such specialisations
- Populate the platform with Wittgenstein's Brown Book and related material
- Have ~10 scholars work in that environment
- Analyse and model the resulting scholarly semantic graph

Expected WP3 Results



- Prototype **platform enabling digital scholarship** in combining EDM RDF metadata, digital surrogates and Linked Data ...
- ... building on an **ontological representation of scholarly work** based on a common understanding of its constituents
- ... resulting in a increasingly complex **social semantic scholarly graph** containing RDF statements such as
 - Version_A – isSuccessorOf – Version_B
 - Statement_1 – contradicts - Statement_2
 - Scribe_Y – copiedFrom – Scribe_Z
- ... and which could **feed back richly contextualised EDM** to Europeana!



Suggested Reading

Tobias Blanke, Mark Hedges (2011): Scholarly primitives. Building institutional infrastructure for humanities e-Science, Future Generation Computer Systems, In Press (Science) (<http://bit.ly/M4AP6w>)

Gregory Crane (2006): What Do you Do with a Million Books? In: Dlib Magazine, Vol. 12, March. (<http://bit.ly/JhzF90>)

Gradmann, S. (2010). Knowledge = Information in Context: on the Importance of Semantic Contextualisation in Europeana. (<http://scr.bi/LAfVzL>)

David Parry: Burn the Boats/Books. Presentation to Digital Writing and Research Lab, Austin. (<http://bit.ly/JYLIJV>)

John Unsworth (2000): Scholarly Primitives. What methods do humanities researchers have in common, and how might our tools reflect this? In the seminar on Humanities Computing, King's College, London. (<http://bit.ly/M4AWz7>)

Thank you for your patience and attention

